Teaching and Learning in Nursing 11 (2016) 133–137

Contents lists available at ScienceDirect

Teaching and Learning in Nursing

journal homepage: www.jtln.org

Towards an Evidence-Based Methodology for High-Stakes Evaluation of Nursing Students' Clinical Performance Using Simulation

Marilyn H. Oermann, PhD, RN, ANEF, FAAN [a,*], Suzan Kardong-Edgren, PhD, RN, CHSE, ANEF, FAAN [b], Mary Anne Rizzolo, EdD, RN, ANEF, FAAN [c]

[a] Duke University School of Nursing, Durham, NC
[b] School of Nursing and Health Sciences, Robert Morris University, Moon Township, PA
[c] National League for Nursing, Washington, DC

ARTICLE INFO

Keywords:
Clinical performance
High-stakes evaluation
Nursing students
Simulation-based assessment

ABSTRACT

The aim of this study was to explore the feasibility of developing scenarios for high-stakes evaluation of students' clinical performance. We used video recordings of graduating nursing students performing in these scenarios. This article shares the findings of the study and challenges we encountered with using simulation for high-stakes evaluation.

Introduction

More nursing programs are beginning to explore the use of simulation as a summative clinical evaluation. There are multiple reasons for seeking new strategies for evaluating clinical performance in addition to the traditional method of observing students providing care and rating their competency. Observations of performance in the clinical setting depend on the teacher being available and present to observe the students, which is difficult in many busy health care agencies. There may be wide variation among clinical teachers in the data collected in an observation of students in the clinical setting and in their judgment of the quality of the students' performance (Oermann & Gaberson, 2014). The number of students for whom the teacher is responsible, the patient population, and other factors also may influence the clinical evaluation process.

Some clinical instructors have limited preparation for assessing student competencies. This lack of preparation may affect the quality of the evaluation and result in reluctance to fail students who do not perform well clinically but are still able to perform well on examinations and written papers (Hunt, McGee, Gutteridge, & Hughes, 2012; Luhanga, Yonge, & Myrick, 2008). In addition, faculty frequently give students "the benefit of the doubt"' when they perform poorly during a clinical course, sometimes passing students who will continue to struggle as they progress through the program (Docherty & Dieckmann, 2015; Wolf et al., 2011).

Nursing programs have traditionally relied on end-of-course tests, clinical evaluation summaries, and commercially prepared tests that mimic the National Council Licensure Examinations (NCLEX) for assessment of nursing students. A high-stakes simulation experience may provide an additional method of evaluating nursing students' clinical performance. The development of simulation as a pedagogy and its wide use in nursing education provide an opportunity to explore using simulation as an assessment method in addition to the current reliance on tests, written assignments, ratings of students' performance in the clinical setting, and NCLEX readiness tests for validating students' knowledge and competencies in nursing. However, before using simulation modalities for high-stakes evaluations of students' clinical performance, more studies are needed to establish the processes and standards for conducting these evaluations.

Purpose and Sample

The aim of this study was to explore the feasibility of developing a series of parallel scenarios for high-stakes evaluation of students' clinical performance. Research findings indicate that rater bias

Funding: The project was coordinated through the National League for Nursing with funding from Laerdal Medical.

Disclosure: The authors acknowledge the following potential conflict of interest: The project was coordinated through the National League for Nursing (NLN) with funding from Laerdal Medical. One of the authors (Rizzolo) is a consultant to the NLN. Other than Dr. Rizzolo, no one from the NLN or Laerdal Medical was involved in the project referred to in the article, in writing the article, or in the decision to submit the article for publication. The authors had sole responsibility for the project and article, and there were no restrictions by the NLN or Laerdal Medical.

* Corresponding author. Tel.: +1 919 684 1623; fax: +1 919 681 8899.
E-mail address: marilyn.oermann@duke.edu.

http://dx.doi.org/10.1016/j.teln.2016.04.001
1557-3087/

(subconscious or conscious) exists when a rater has worked with a student in the classroom or clinical setting (Oermann & Gaberson, 2014; Stroud, Herold, Tomlinson, & Cavalcanti, 2011). In many nursing programs, it might be difficult to find a faculty member who is not familiar with a student. Thus, in this study, as a methodology to ameliorate potential bias in a high-stakes assessment, we used video recordings of students performing in standardized scenarios. Previous research has found no differences in scoring between ratings done in real time and those of recorded scenarios (Williams et al., 2009).

The sample included volunteer graduating nursing students from varied types of prelicensure programs across the United States. In this article, we report the findings of the pilot and field tests using this methodology, describe our experiences with simulation for high-stakes evaluation, and share some solutions to challenges we identified.

Outcomes Evaluated

Four outcomes were chosen for the evaluation of students' clinical performance in a simulation based on the results of an expert panel think tank: (a) assessment and intervention, (b) nursing judgment, (c) quality and safety, and (d) teamwork and collaboration, with communication integrated across the other areas. The think tank was convened by the National League for Nursing, with the goal of exploring the feasibility of using manikin-based high-fidelity simulation for evaluating students' performance at the end of a prelicensure nursing program. Invited participants at the think tank included experts in curriculum design, competency-based evaluation, clinical evaluation, tool development, and simulation in both nursing and medical education (Rizzolo, Kardong-Edgren, Oermann, & Jeffries, 2015). As part of their work, they reviewed sample program outcomes from prelicensure nursing programs throughout the United States and selected outcomes that could be assessed using simulation.

Our goal was to develop scenarios for evaluating students' performance related to each of the four outcomes (assessment and intervention, clinical judgment, quality and safety, and teamwork and collaboration). However, we realized that "quality" would be difficult to assess in a scenario and that we could not evaluate "teamwork" because the scenarios were designed to be run with a single student participant. High-stakes evaluation was defined as an evaluation process associated with a simulation activity that has a major academic or educational consequence, such as a grading decision, including pass or fail implications, or a decision regarding competency. High stakes refers to the outcome or consequences of the process (Meakim et al., 2013).

Scenario Development and Pilot Test

Each scenario was written with three parallel but equal variations to lessen the impact of students sharing information about the scenario with peers (Willhaus, Burleson, Palaganas, & Jeffries, 2014). The simulation experts who developed the scenarios identified the key behaviors expected of students for their successful completion. We provided the following information to the participating schools: the outcome evaluated with a description of it; the key behaviors expected of students in the scenario; a description of the case; needed equipment and simulation preparation, including preparation of the environment, manikin, equipment, medications, and fluids; documentation and scenario forms; and other information about the simulation. A standard template was used for each scenario including time, monitor settings, patient/monitor actions, student intervention, and patient/additional participant responses. Students received a verbal patient report before beginning, which we provided to the schools.

Scenarios were planned to be run with one student performing the scenario over 20–30 minutes. This time included an initial 5–10 minutes for students to review the chart and orient themselves to the manikin and room. A formal debriefing was not included as the scenarios were meant to be summative. The role of summative evaluation is to determine if the student has achieved the outcomes and can perform competently, not for the teacher to give feedback to guide future learning such as in a debriefing (Oermann & Gaberson, 2014). Debriefing information also might be shared by students with others in the school, decreasing the reliability of the scenarios.

All 12 simulations (three parallel versions of each of the four scenarios) were piloted at 10 schools of nursing and video recorded. These schools included one diploma, four associate degree, and five baccalaureate programs. All regions of the United States were represented except for Western states; there were two schools from the North Atlantic, two from the Midwest, and five from the South. None of the evaluators were faculty members at these schools. The videos were deidentified and numbered for use in the study. The clinical judgment scenario was pilot tested first. The authors realized when viewing the videos of that scenario that evaluators could not observe and assess students' performance of psychomotor skills such as preparing intravenous (IV) medications. The cameras used in the schools did not allow as much zoom capability as was required for that level of detail.

In the pilot videos, the level of cueing of students varied from school to school. Unexpected changes made to the script of the verbal shift reports, changes of intonation, and the emphasis placed on portions of the verbal report by some simulation facilitators led students down different paths of decision making. Better scripts, more explicit instructions to the schools of nursing, and a prerecorded shift report solved some of these problems in the next phase of the study. We also added instructions about overcueing students and the intended length of the scenarios. We learned that the conversations with off-camera confederates, such as physicians on the phone, need to be clearly heard in the video recordings for the evaluators to judge the student's communication skills.

Evaluation Tool

A checklist style scoring system was considered initially for scoring student performance in the scenarios but was discarded. Checklists capture lower level skill performance but not the higher order thinking skills required for clinical practice (Boulet 2008; Boulet et al., 2008; Oermann & Gaberson, 2014). Checklists also promote memorization and may limit critical thinking. For these reasons, we decided not to use checklists for the evaluation of performance in the scenarios. Multiple tools were evaluated, and the Creighton Competency Evaluation Instrument (CCEI; Hayden, Keegan, Kardong-Edgren, & Smiley, 2014), a modification of the Creighton Simulation Evaluation Instrument (Parsons et al., 2012; Todd, Manz, Hawkins, Parsons, & Hercinger, 2008), was chosen for the study.

The CCEI includes 23 skills grouped into four main areas similar to the outcomes evaluated in the study. This tool was used for the National Council of State Boards of Nursing National Simulation Study for evaluating students' clinical competencies in both simulation and traditional clinical experiences. Validity and reliability are both high when raters are trained appropriately. Hayden et al. (2014) reported that content validity ranged from 3.78 to 3.89 on a 4-point Likert scale. Cronbach's alpha was >.90 when used to score three different levels of simulation performance.

Field Test Evaluator Selection and Training

After completion of the pilot study, 10 faculty members from associate and baccalaureate programs were selected to evaluate the